Systematic Stepsize Variation: Efficient Method for Searching Conformational Space of Polypeptides

CHRISTIAN T. KLEIN, BERND MAYER, GOTTFRIED KÖHLER, PETER WOLSCHANN

Institute for Theoretical Chemistry and Radiation Chemistry, University of Vienna, Währinger Strasse 17, A-1090 Vienna, Austria

Received 12 January 1998; accepted 20 April 1998

ABSTRACT: A new and efficient method for overcoming the multiple minima problem of polypeptides, the systematic stepsize variation (SSV) method, is presented. The SSV is based on the assumption that energy barriers can be passed over by sufficiently large rotations about rotatable bonds: randomly chosen dihedral angles are updated starting with a small stepsize (i.e., magnitude of rotation). A new structure is accepted only if it possesses a lower energy than the precedent one. Local minima are passed over by increasing the stepsize systematically. When no new structures are found any longer, the simulation is continued with the starting structure, but other trajectories will be followed due to the random order in updating the torsional angles. First, the method is tested with Met-enkephalin, a peptide with a known global minimum structure; in all runs the latter is found at least once. The global minimum conformations obtained in the simulations show deviations of ± 0.0004 kcal/mol from the reference structure and, consequently, are perfectly superposable. For comparison, Metropolis Monte Carlo simulated annealing (MMC-SA) is performed. To estimate the efficiency of the algorithm depending on the complexity of the optimization problem, homopolymers of Ala and Gly of different lengths are simulated, with both the SSV and the MMC-SA method. The comparative simulations clearly reveal the higher efficiency of SSV compared with MMC-SA. © 1998 John Wiley & Sons, Inc. J Comput Chem 19: 1470–1481, 1998

Keywords: protein folding; simulated annealing; Monte Carlo dynamics; multiple minima problem; global minimum conformation

Correspondence to: C. T. Klein; e-mail: christian.klein@tbi.univie.ac.at

Contract/grant sponsor: Austrian Academy of Sciences

CCC 0192-8651 / 98 / 131470-12

Introduction

The large number of minima on the energy hypersurface of a protein arises from the fact that a polypeptide consisting of n residues possesses at least $2n$ freely rotatable bonds (the backbone dihedrals). In force fields developed for the simulation biomolecules, such as AMBER[1] or GROMOS,[2] the corresponding torsional angles, Φ_i, are described by sinusoidal potentials of the form:

$$V(\Phi_i) = k_i[1 + \cos(n_i\Phi_i - \delta_i)]$$

where k_i are force constants, n_i are the multiplicities of the dihedrals (identical to the number of minima of the potential function), and δ_i are the phases. The empirical conformation energy program for polypeptides (ECEPP) force field[3–6] also uses a potential of the form just given for calculation of the energy contributions stemming from side-chain torsional angles.

In the GROMOS force field[2] the multiplicity for both ϕ and ψ backbone dihedrals is 6, which leads to a number of minima for the n residue peptide of the order 6^{2n}.[7] From these considerations, it follows that the number of possible minima exponentially increases to astronomic values, even though many of them are not accessible, due to steric hindrance. Therefore, efficient methods for searching the conformational space are necessary to find representative local minima, or even the global minimum, of a given polypeptide.

In recent years, a variety of methods have been developed to overcome the multiple minima problem.

In *molecular dynamics* (MD) simulations, energy barriers are passed over due to the kinetic energy of the molecular system. Unfortunately, the timesteps used in such simulations are very small. Therefore, the gap between the time scale of the simulation, which for large proteins is limited to the order of hundreds of picoseconds, and that of protein folding, which is of the order of magnitude from milliseconds to minutes, makes MD unsuitable for global minimization. However, constrained molecular dynamics simulations have been applied successfully in folding small proteins from the extended to the native conformation.[8]

Stochastic methods are usually more efficient in sampling the conformational space of polypeptides. The *Metropolis Monte Carlo method* (MMC)[9] generates a canonical ensemble of structures, among which low energy structures or even the global minimum structure may be found: The molecule is stochastically deformed and the new structure is accepted or rejected by the Metropolis criterion; that is, with a probability given by the Boltzmann factor. In this way energy barriers can be passed over.

Combining MMC with energy minimization (*Monte Carlo minimization*)[10] increases its efficiency, as only local minima are compared by the MMC criterion.

Because the MMC method is temperature-dependent, it can be combined with simulated annealing[11,12] (MMC-SA). Within MMC-SA simulations, the temperature is lowered (stepwise or exponentially) during the run. At higher temperatures energy barriers can be overcome more easily, and thus a wider range of the conformational space is sampled. As the temperature decreases, the search is more and more restricted to smaller regions of the energy hypersurface. With optimal cooling schedules, the probability of finding the global minimum of a structure is much higher than with conventional MMC. However, a major problem of simulated annealing applied to polypeptides is the fact that the optimal cooling schedule cannot be realized in most cases of interest. This problem can be eliminated by using annealing methods that evolve in temperature rather than in time. One of these methods is Gaussian density annealing (GDA).[13] The equations of motion, derived from the classical Bloch equation, are integrated in (reciprocal) temperature, starting from a known infinite temperature distribution to a final value, $\beta_{\max}$.

Monte Carlo methods at fixed temperature that use acceptance criteria other than the Metropolis criterion are also successfully used in overcoming the multiple minima problem of polypeptides. Because MMC adds weight to the configurations with a Boltzmann factor, the probability distribution, $P_B(E)$, of the energies (E) has a bell-like shape, and its values for low E are smaller by many orders of magnitude than the maximum value of $P_B(E)$. In multicanonical algorithms,[14] the probability distribution is defined in such a way that a configuration of any energy is accepted with equal probability. Thus, the global minimum energy region of a polypeptide is explored in a much shorter simulation time (within a given statistical quality) compared with the canonical Monte Carlo method at fixed temperature. In the enhanced

sampling algorithm approach,[15] a generalized statistical distribution is used, resulting in an increased rate of barrier crossing.

A class of stochastic methods used more and more in searching the conformational space of polypeptides are based on genetic algorithms.[16] Populations of individuals (i.e., conformations) are generated. The ''survival chance'' of an individual depends on its ''fitness score'' (i.e., on the energy of the respective conformation). New populations can evolve from the old ones by mutation and selection. Genetic algorithms are efficient optimization strategies, especially for smaller molecules.[17] An efficient method that is a variant of a genetic algorithm, conformational space annealing, has recently been published.[18]

A very promising class of methods is based on smoothing the energy hypersurface, such that, finally, all minima except the global minimum disappear. Among these, the diffusion equation method (DEM)[19] has been successfully used in finding the global minimum of small peptides[20] and of a 36-residue C-terminal domain of an enzyme.[21]

In this work, we present a simple and efficient method for sampling the conformational space of polypeptides. First, the method is tested with a 24-dihedral-angle problem, namely the well-studied pentapeptide, Met-enkephalin (H_2N—Tyr—Gly—Gly—Phe—Met—COOH). Considering the multiplicities of the torsional angle potentials, it is assumed to possess more than 3^{24} (10^{11}) local minima.[10] Because the accepted global minimum structure has been reported by a number of groups,[6, 10, 18, 20, 22, 23] this peptide is often used to test the efficiency of algorithms developed for overcoming the multiple minima problem. To get results comparable to those reported,[6, 18] the peptide is studied with the ECEPP/3 force field.[6]

The dependence of efficiency on the magnitude of the problem is then analyzed by using model peptides of different lengths; that is, homopolymers of an α-helix former (Ala) and an α-helix breaker (Gly): H_2N–$(Ala/Gly)_n$—COOH, with $n = 6, 8, 10, 12, 14, 16$. For these peptides, the GROMOS-87 force field[2] is used, for which, as already mentioned, the number of local minima increases rapidly with 6^m, where m is the number of rotatable bonds.

To estimate the efficiency of the method, it is compared with a ''classical'' optimization method: Metropolis Monte Carlo simulated annealing (MMC-SA).[11, 12]

For obvious reasons, stemming from the algorithm, the method will subsequently be called the systematic stepsize variation (SSV) method.

SSV Method

The basic idea of the SSV method for sampling the conformational space is that energy barriers may be passed over by sufficiently large rotations about rotatable bonds. In what follows, a stepsize will denote the magnitude of such a rotation.

Consider a starting structure with bond lengths and bond angles set to their ideal values. In this case, the potential energy of the system depends only on the set of l torsional angles:

$$S_0^{(0)} = \{\alpha_1^{(0)}, \alpha_2^{(0)}, \ldots, \alpha_l^{(0)}\}$$

To reduce the size of the optimization problem, only the $n (n < l)$ rotatable bonds will be treated as variables. The potential energy of the molecular system is thus $E(S_0^{(0)})$, with:

$$S_0^{(0)} = \{\alpha_1^{(0)}, \alpha_2^{(0)}, \ldots, \alpha_n^{(0)}\}$$

Within one update cycle, n updates of the rotatable torsional angles are performed as follows:

1. A randomly chosen dihedral angle is updated by a small but fixed value, $\Delta\alpha$, called initial stepsize, $\alpha_i^{(1)} = \alpha_i^{(0)} + \Delta\alpha$. The new structure is defined by:

   $$S_1^{(1)} = \{\alpha_1^{(0)}, \alpha_2^{(0)}, \ldots, \alpha_i^{(1)}, \ldots, \alpha_n^{(0)}\}$$

   The notation, $S_j^{(k)}$, refers to the kth individual step (update) in the jth update cycle.
2. The new structure is accepted if its energy is lower than that of the precedent one; that is, $E(S_1^{(1)}) < E(S_0^{(0)})$, and another randomly chosen dihedral angle is updated.
3. In the contrary case, $\alpha_i^{(0)}$ is updated by substituting it with $\alpha_i^{(1)} \rightarrow \alpha_i^{(0)} - \Delta\alpha$.
4. If, again, the new structure possesses a lower energy than the precedent one [i.e., $E(S_1^{(2)}) < E(S_1^{(1)})$], it will be accepted and another randomly chosen torsional angle will be updated.
5. If none of the two substitutions have been accepted, $\alpha_i^{(0)}$ is maintained unchanged and another stochastically chosen dihedral angle is updated.

In one full update cycle, steps 1–5 are performed n times. For a better understanding, one update cycle is presented as a flow chart in Figure 1.

6. If, in an update cycle, at least one structure with a lower energy than the precedent cycle is found, a new cycle is started with the same initial stepsize, $\Delta\alpha$. In the opposite case, the system is trapped at a minimum. To overcome the energy barriers surrounding the minimum, in the next cycle the stepsize is increased by $\Delta\alpha'$, called update stepsize.
7. If, with the new stepsize $\Delta\alpha + \Delta\alpha'$, at least one structure with lower energy is found within a cycle, the stepsize is reset to the initial value of $\Delta\alpha$, otherwise it is increased again by $\Delta\alpha'$ in the next cycle.
8. Steps 6–7 are repeated until stepsize $\Delta\alpha + i\,\Delta\alpha'$ has reached 360°. In this situation, the structure is either at the global minimum, or at a local minimum that cannot be left by changing only one torsional angle per step, and cannot be further optimized by employing steps 1–7. Therefore, the simulation is restarted with the initial geometry $S_0^{(0)}$; however, because the dihedrals are updated in random order, a different trajectory will be followed, and other minimum structures will be found.

The flow chart of steps 6–8 is presented in Figure 2.

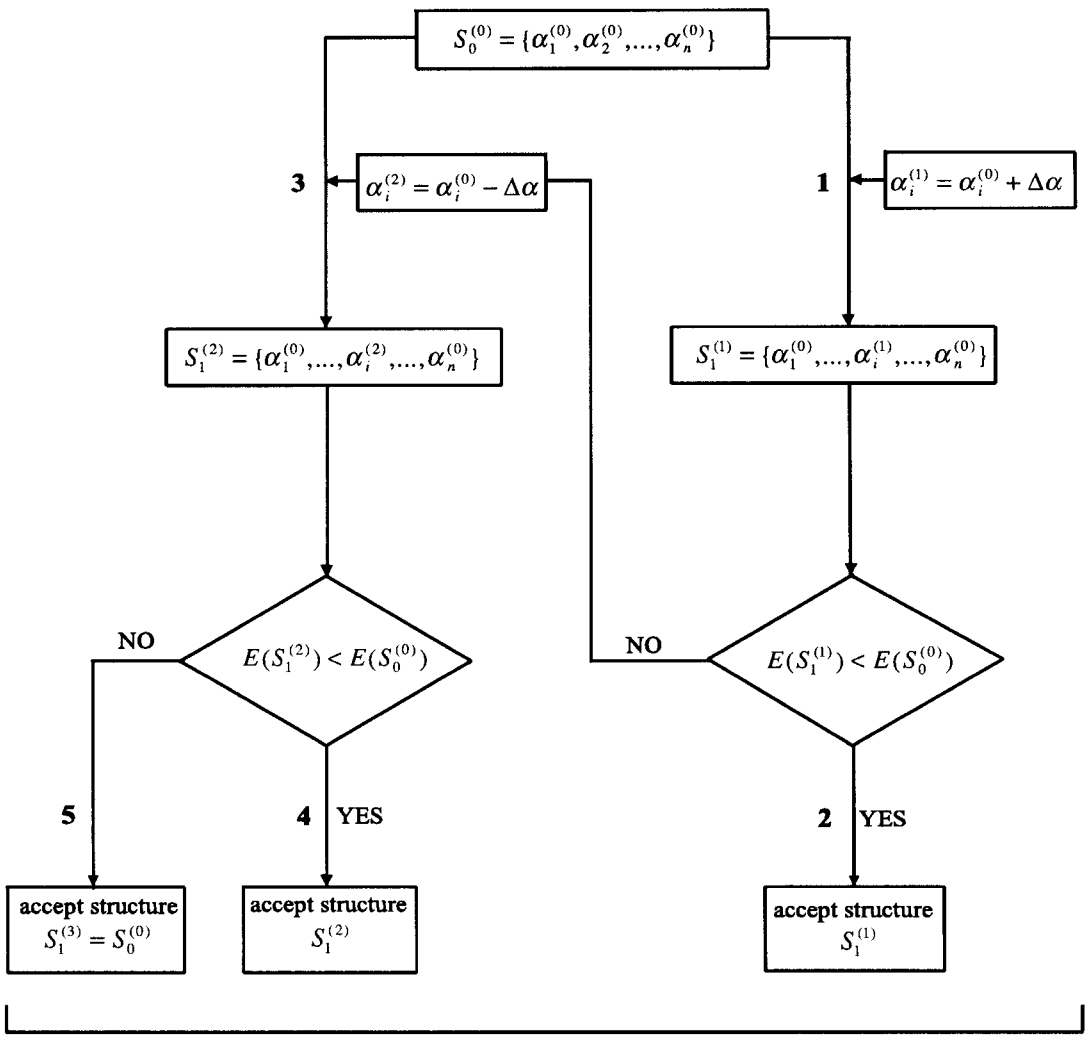

FIGURE 1. Flowchart of one update cycle with the SSV method.

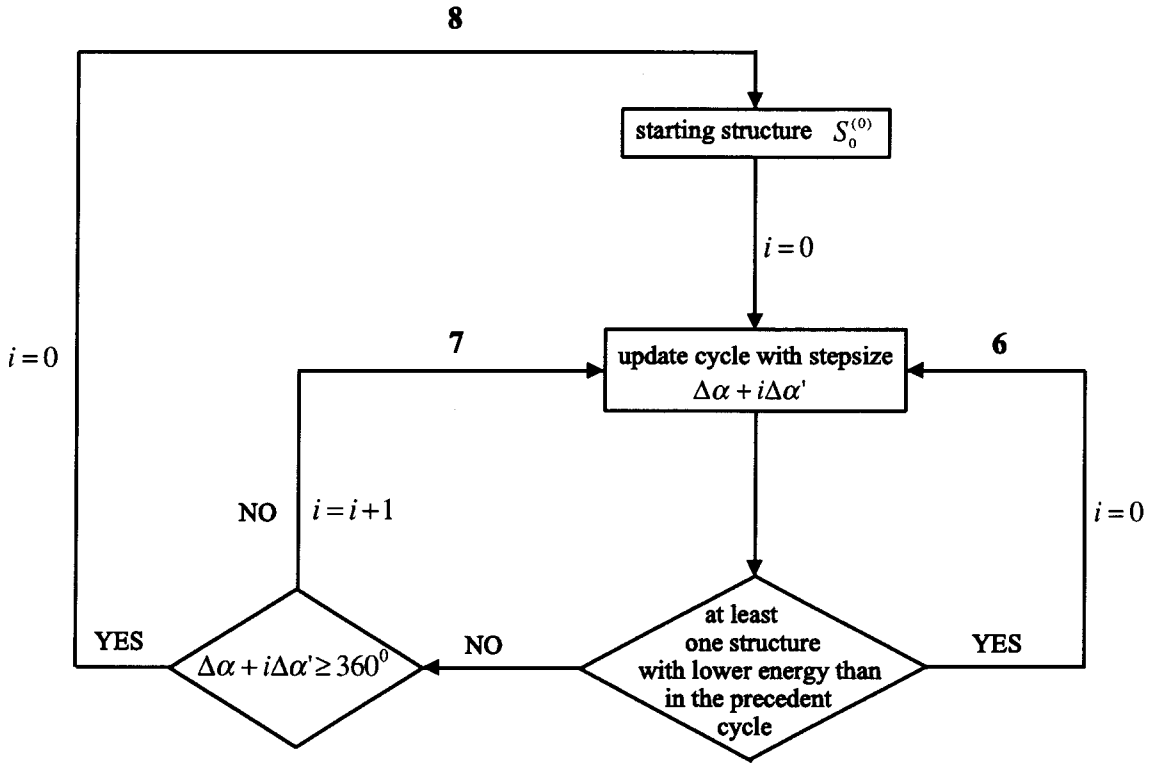

FIGURE 2. Flowchart of the stepsize variation within the SSV algorithm.

METROPOLIS MONTE CARLO SIMULATED ANNEALING (MMC-SA)

In the MMC-SA[12] simulations used for comparison, structures are generated by randomly choosing a torsional angle and stochastically updating it in the range $-180°$ to 180°. The new structure is accepted or rejected with a probability given by the Boltzmann factor:

$$P(\Delta E, T) = \min\left[1, \exp\left(-\frac{\Delta E}{RT}\right)\right]$$

where ΔE is the energy difference between a new and an old structure, T the absolute temperature, and R the gas constant. The temperature of the system is cooled down exponentially from 1000 K to 250 K during one annealing cycle.

Computational Details

In the simulations, the GROMOS-87 force field,[2] with a distance-dependent dielectricity constant[24] and the ECEPP/3 force field,[6] are used.

To ensure comparability to published results,[6, 10, 18] the ECEPP/3 force field is used with the SUMSL minimizer[25] option for Met-enkephalin; that is, in each step the structure is fully minimized. Therefore, a large initial stepsize of $\Delta\alpha = 30°$ and an updated stepsize of $\Delta\alpha' = 60°$ have been shown to be optimal for exploring the conformational space. If the initial stepsize is too small, the system falls back to the same minimum after energy minimization. If, for example, stepsizes of $\Delta\alpha = 15°$ and $\Delta\alpha' = 60°$ are used, the frequency of finding the global minimum conformation of Met-enkephalin is reduced by half. The same is true if the initial stepsize is too large—that is, for stepsizes of $\Delta\alpha = 60°$ and $\Delta\alpha' = 60°$.

All dihedrals, including the ω angles, are treated as variables. Fifteen runs are performed with both the SSV and the MMC-SA methods. Each run consists of 30,000 steps, which, in the case of MMC-SA, corresponds to three complete simulated annealing cycles of 10,000 steps. Also, other cooling schedules have been tested, for example, with 6000 steps per cycle, but the 10,000-step schedule has proven to be more efficient.

For the homopolymers simulated with the GROMOS force field, the ω dihedrals are fixed to their ideal values of 180°. The structures are not fully minimized, but only slightly relaxed within one minimization step. In this case, the initial stepsize should be as small as possible to ensure

an efficient sampling of the local minima. The updated stepsize should be large enough to permit a rapid overcoming of energy barriers, but small enough to still ensure a thorough exploration of the energy hypersurface. Values of $\Delta\alpha = 5°$ and $\Delta\alpha' = 5°$ were found to give optimum results when searching the conformational space in the case of the homopolymers.

The simulations with both the SSV method and MMC-SA consist of the same number of steps per run: 90,000 for $(Ala/Gly)_6$ and $(Ala/Gly)_8$ and 200,000 for $(Ala/Gly)_{10}$, $(Ala/Gly)_{12}$, $(Ala/Gly)_{14}$, and $(Ala/Gly)_{16}$. For MMC-SA, the runs of 90,000 steps and correspond to 15 complete simulated annealing cycles of 6000 steps, and those of 200,000 steps to 20 complete cycles of 10,000 steps per cycle.

The starting structures in all simulations are the completely extended, fully minimized conformations.

Results and Discussion

MET-ENKEPHALIN

In all of the 15 runs, the accepted global minimum conformation[6, 18] is found several times with the SSV method. For each run, Figure 3 shows the stepnumber at which a conformation with an energy below -11.706 kcal/mol is found. The global minimum is found 156 times in 450,000 steps with the SSV method; that is, on average, every 2885th step (after 2885 energy minimizations).

Comparing the lowest energy structures to the reference global minimum (-11.7068 kcal/mol)[18] reveals that they are identical, with energies between -11.7062 kcal/mol and -11.7072 kcal/mol. In Table I, five of these structures are presented together with the reference structure.

At a first look the conformations seem to be different from the reference structure. The dihedrals that differ from the reference are printed in bold in Table I. Nevertheless, a more precise analysis shows that all six conformations from the table are identical; consider, for instance, the χ^2 angle of phenylalanine. In the reference structure the χ^2 angle has a value of 94.5°, in structures 1–4 it has values of $-85.4°$. The difference is a phase shift of exactly 180°; that is, the phenyl ring is in the same position for both values. The same argument is valid for the χ^4 angle of methionine, in which the different values of $-178.6°$, $-58.5°$, and 61.4 are shifted by 120°, corresponding to the same orientation of the methyl group.

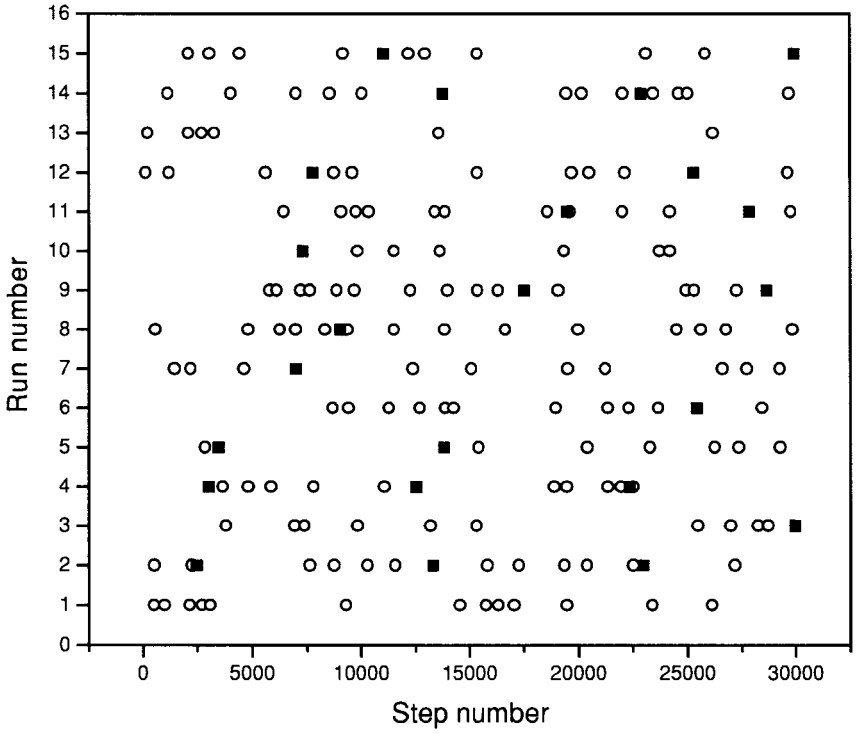

FIGURE 3. Step number at which structures with energies below -11.706 kcal / mol are found in the 15 runs. The open circles denote the SSV, the solid squares the MMC-SA method.

TABLE I.
Five Global Minimum Conformations Obtained in Simulations With SSV.[a]

	Reference	1	2	3	4	5
Try						
ϕ	−83.499	−83.473	−83.470	−83.393	−83.352	−83.431
ψ	155.901	155.812	155.803	155.807	155.765	155.805
ω	−177.181	−177.121	−177.119	−177.154	−177.120	−177.143
χ^1	−173.198	−173.174	−173.183	−173.188	−173.191	−173.158
χ^2	79.366	79.306	79.323	**−100.637**	**−100.612**	**−100.673**
χ^3	−166.362	−166.327	−166.337	**13.641**	**13.660**	**13.652**
Gly						
ϕ	−154.305	−154.284	−154.276	−154.323	−154.258	−154.310
ψ	86.000	85.818	85.822	85.950	85.852	85.817
ω	168.502	168.516	168.512	168.515	168.505	168.504
Gly						
ϕ	82.920	82.968	82.959	82.948	82.986	83.014
ψ	−75.096	−75.038	−75.024	−75.090	−75.039	−75.060
ω	−169.977	−169.974	−169.966	−169.964	−169.974	−169.981
Phe						
ϕ	−136.889	−136.859	−136.856	−136.856	−136.863	−136.818
ψ	19.116	19.113	19.085	19.091	19.098	19.087
ω	−174.093	−174.100	−174.073	−174.088	−174.110	−174.071
χ^1	58.865	58.856	58.867	58.868	58.860	58.854
χ^2	94.545	**−85.468**	**−85.479**	**−85.483**	**−85.465**	94.520
Met						
ϕ	−163.486	−163.454	−163.445	−163.442	−163.434	−163.436
ψ	161.184	160.940	160.948	160.958	160.956	160.893
ω	−179.799	−179.785	−179.805	−179.784	−179.813	−179.764
χ^1	52.889	52.966	52.868	52.876	52.859	52.869
χ^2	175.290	175.303	175.264	175.309	175.309	175.291
χ^3	−179.854	−179.872	−179.848	−179.863	−179.864	−179.860
χ^4	−58.545	**61.413**	**−178.580**	**61.402**	**−178.591**	**−178.59**
Energy (kcal / mol)	−11.7068	−11.7072	−11.7072	−11.7065	−11.7064	−11.7062

[a] All structures are identical to the reference conformation (taken from ref. 18). Dihedrals that are apparently different from the reference structure are indicated in bold.

In Figure 4, all six structures from Table I are overlayed: they are perfectly superposable (i.e., they show no visible rms deviations). The backbones of the polypeptides are drawn with a thick line. Figure 4 shows that, at the global minimum, the backbone of Met-enkephalin forms a loop that is flanked by the two phenyl rings of the aromatic amino acids, Tyr and Phe, respectively.

With MMC-SA the global minimum conformation is found less frequently. The 15 runs of 30,000 steps, presented in Figure 3, correspond to 45 complete simulated annealing cycles of 10,000 steps each. The global minimum is found in 23 cycles, which means, on average, every 19,565th simulation step.

Because the trajectories go only ''downhill'' when the SSV algorithm is employed, the number of accepted structures is smaller, compared with MMC-SA, within the same number of steps. In all runs (450,000 steps for both methods) with SSV 11,346 structures are accepted, whereas with MMC-SA the number of accepted structures is 53,732 (i.e., about five times more). Nevertheless, the percentage of low energy structures found is

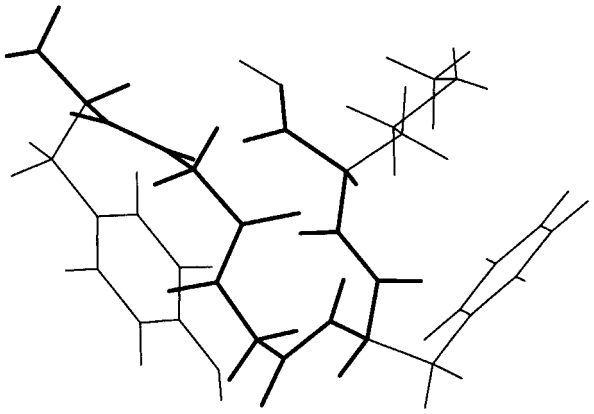

FIGURE 4. Superposition of the five global minimum structures from Table I, obtained with the SSV method, together with the reference structure (from ref. 18).

higher for SSV: 29.3% of the conformations have energies below −10 kcal/mol, whereas for MMC-SA it is only 7.2%.

In Figure 5, a histogram of the number of structures found in the energy intervals from −12.0 to −11.0 to −6.0 to −5.0 kcal/mol, is presented. For MMC-SA, the distribution has, as expected, a bell-like shape, with a maximum (9782 structures) between −8.0 and −7.0 kcal/mol. For SSV, there are two, almost equal maximum peaks from −11.0 to −10.0 kcal/mol (2880 structures) and −6.0 to −5.0 kcal/mol (2937 structures). This situation reflects the nature of the SSV algorithm: The simulation starts from the extended conformation, which is, according to the Ramachandran diagram,[26] close to an allowed region of (ϕ, ψ) space, namely that of β-sheets. Therefore, acceptance at the beginning will be relatively high in this region of the conformational space. After it has been explored, the system moves rapidly, upon successively increasing the update stepsize, to low energy regions of the hypersurface, which is sampled until a ''dead end'' structure is found. This corresponds to a minimum that cannot be overcome by the algorithm and a new trajectory is started from the extended conformation.

Almost all of the low energy conformations found in this range are found by both methods. In Table II, conformations with $E < -10$ kcal/mol are listed together with their distance matrix errors (DMEs) relative to the global minimum conformation (−11.707 kcal/mol). The method (MMC-SA or SSV) by which the respective structure is found is also indicated in Table II.

The efficiency of SSV can also be pointed out by comparing it with a recently published method, conformational space annealing (CSA),[18] which is a variant of a genetic algorithm. The conformational space is divided by a cutoff distance into subdivisions of local minima. Conformations within the same cutoff are considered to belong to the same local minimum region and represent a population. At the beginning the cutoff is set to large values, such that the whole conformational space is covered. Annealing is achieved by slowly reducing the cutoff and replacing groups of high energy local minima with groups of low energy local minima. In this way the method finds not only the global minimum energy conformation of Met-enkephalin, but also distinct local minima as byproducts. The frequency of finding the global minimum conformation with the two methods is comparable: with SSV it is found, on average, every 2885th step

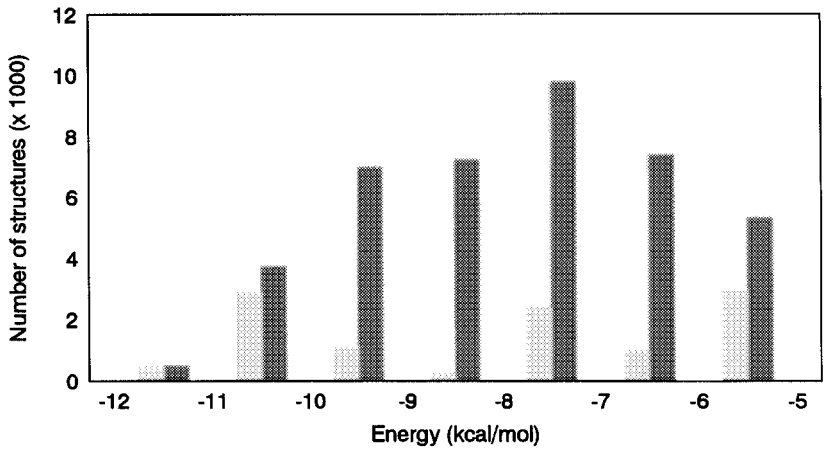

FIGURE 5. Histogram of the number of structures found in the interval −12.0 to −11.0 kcal / mol to −6.0 to −5.0 kcal / mol. The SSV method is denoted by light gray bars, MMC-SA by dark gray bars.

TABLE II.
Structures of Met-Enkephalin with Energies Below −10 kcal / mol.

Energy (kcal / mol)	DME	SSV	MMC-SA
− 11.707	0.00	+	+
−11.229	0.65	+	+
−10.980	0.53	+	+
−10.964	1.37	+	+
−10.958	0.82	+	+
−10.856	0.71	+	+
−10.552	1.18	+	−
−10.547	2.29	−	+
−10.449	1.46	−	+
−10.432	0.77	+	+
−10.384	0.75	+	+
−10.346	1.38	+	+
−10.334	0.70	+	−
−10.279	1.15	+	+
−10.242	1.60	+	+
−10.221	1.58	+	+
−10.195	1.08	+	+
−10.151	1.32	+	+
−10.147	1.29	+	+
−10.132	1.11	+	+
−10.123	1.25	+	+
−10.068	1.15	+	+
−10.055	1.23	−	+
−10.037	1.57	+	−
−10.023	1.24	+	+

+ indicates whether the conformation is found with the SSV or the MMC-SA method, respectively.

(i.e., after every 2885th minimization—which is the most time-consuming part of the optimization step), whereas, with CSA it is found, on average, after every 2600th energy minimization. An analysis of the torsional angles shows that the seven lowest energy conformations found with CSA are identical to the seven lowest energy structures found with SSV (the first seven listed in Table II).

HOMOPOLYMERS

In the case of the polyalanines the lowest energy structures are always α-helices extending over the whole polypeptide. We define the maximal helix length, l_{max}, as the maximal number of successive helical residues. A residue is considered helical if its ϕ- and ψ-dihedrals do not differ by more than 30° from the ideal (ϕ, ψ)-values of an α-helix $(-58°, -47°)$.

Table III presents the results of the simulations of the SSV method as compared with MMC-SA. Five runs are performed for each homopolymer. The lowest energies found within the same number of steps [90,000 for H_2N—$(Ala)_n$—COOH ($n = 6, 8$) and 200,000 for H_2N—$(Ala)_n$—COOH ($n = 10, 12, 14, 16$)] are shown. With both the SSV and MMC-SA methods, the lowest energy structures found in each run are α-helices extending over the whole molecule for the polyalanines up to 10 residues. Nevertheless, the conformations found with SSV have lower energies, although the distance matrix errors (DMEs) of the ideal α-helix show fluctuations within the same limits for both methods. The average differences between SSV and MMC-SA are about −1 kcal/mol for $n = 6$, −3 kcal/mol for $n = 8$, and −4 kcal/mol for $n = 10$. These differences, corresponding basically to the same DMEs, can be explained by the fact that, for the small initial stepsize of $\Delta\alpha = 5°$, the SSV method resembles a local minimization strategy.

The slight decrease of the DMEs when the number of residues increases from 6 to 10 stems from the fact that, in the ''real'' α-helices, one or both terminal residues are not in the ''helical'' conformation; that is, their (ϕ, ψ)-values differ by more than 30° from the ideal values. Indeed, if the simulation is started from the perfect α-helix of length n ($n = 6, 8, 10, 12, 14, 16$), one always obtain lower energy helices of length $n - 1$ with the SSV method. Consequently, the DMEs of the ideal helices will decrease with an increasing number of residues.

For the polyalanines with 12−16 residues the energy differences between SSV and MMC-SA continue to get larger with an increasing number of residues (local minima): about −5 kcal/mol for $n = 12$; −10 kcal/mol for $n = 14$; and −19 kcal/mol for $n = 16$. For $n = 12$, long helices ($l_{max} = 8, 9, 11$) are still found for both methods, whereas for the longer polymers the situation changes drastically: For $n = 14$, long helices ($l_{max} = 13$) are found in only one of the runs with MMC-SA, whereas in one run a helix of $l_{max} = 9$ and in three runs helices of $l_{max} = 13$ are found with SSV. For $n = 16$, long helices of $l_{max} = 10, 11, 13, 15$ are found only with the SSV method. This is reflected also in the DMEs, which are much larger for SSV when compared with MMC-SA.

The situation is different for the polyglycines. In all simulations, the minimum energies found are, on average, the same for both methods. The results are presented in Table IV.

TABLE III.
SSV Runs Compared with MMC-SA for Polyalanines.[a]

$H_2N—(Ala)_n—COOH$

		SSV			MMC-SA		
n	Nr_{minima}	E_{min} (kcal / mol)	l_{max}	DME	E_{min} (kcal / mol)	l_{max}	DME
6	2.2×10^9	−16.15	5	0.71	−15.23	5	0.74
		−16.28	5	0.76	−15.31	4	0.77
		−15.97	5	0.75	−15.45	4	0.78
		−16.22	5	0.70	−14.84	4	0.76
		−16.39	4	0.80	−14.36	5	0.73
8	2.8×10^{12}	−26.75	7	0.66	−24.23	7	0.67
		−27.01	7	0.69	−24.19	6	0.86
		−27.27	7	0.63	−24.70	7	0.88
		−27.07	7	0.68	−24.83	7	0.61
		−27.23	7	0.65	−24.22	6	0.64
10	3.6×10^{15}	−38.62	9	0.62	−34.83	9	0.57
		−38.03	9	0.57	−34.56	8	0.60
		−38.71	9	0.56	−34.68	9	0.68
		−38.05	9	0.62	−32.96	10	0.61
		−38.04	9	0.62	−34.76	9	0.72
12	4.7×10^{18}	−49.46	11	0.59	−43.92	8	0.85
		−48.27	11	0.91	−41.48	9	0.76
		−47.64	9	0.79	−40.34	11	0.92
		−48.10	9	0.84	−44.91	11	0.75
		−44.70	8	0.78	−43.03	8	0.86
14	6.1×10^{21}	−60.91	13	0.59	−52.46	13	0.57
		−60.74	13	0.46	−51.37	6	3.25
		−55.10	9	2.23	−45.97	5	3.58
		−61.64	13	0.54	−43.65	1	4.13
		−53.18	5	3.21	−46.16	2	4.92
16	7.9×10^{24}	−68.29	13	0.87	−53.67	1	4.95
		−70.27	11	0.63	−42.06	1	4.98
		−68.49	10	3.47	−53.93	2	5.53
		−63.17	5	5.59	−49.87	2	6.40
		−71.69	15	0.48	−46.64	1	4.98

[a] Nr_{minima} denotes the number of estimated local minima,[7] DME the distance matrix error with respect to the ideal α-helix conformation, l_{max} the maximal helix length, and E_{min} the lowest conformational energy found in the respective run.

While for the polyalanines the lowest energy structures are always found to be in the same region of the conformational space, namely that of α-helices, the lowest energy conformations of the homoglycines are spread over a wider range of the conformational space. The DMEs from Table IV are calculated with reference to the lowest energy structure found with SSV or MMC-SA. The values show that structures with approximately the same energies may have DME differences up to a value of 5. This may be explained by the completely different energy hypersurface of the polyglycines as compared with the polyalanines. The lack of side chains in the case of the polyglycines makes minima accessible that would otherwise be forbidden due to steric hindrance.

TABLE IV.
SSV Runs Compared with MMC-SA for Polyglycines.[a]

$H_2N—(Gly)_n—COOH$

		SSV		MMC-SA	
n	Nr_{minima}	E_{min} (kcal / mol)	DME	E_{min} (kcal / mol)	DME
6	2.2×10^9	−9.26	1.76	−9.76	0.59
		−9.49	6.07	−9.72	1.00
		−8.58	1.27	−8.93	1.07
		−10.42	**0.00**	−8.33	1.49
		−9.65	0.77	−9.54	0.66
8	2.8×10^{12}	−15.28	1.99	−15.27	1.33
		−16.41	5.68	−15.15	1.37
		−14.65	1.15	−16.08	0.87
		−16.96	**0.00**	−15.41	1.12
		−14.22	3.23	−14.65	3.39
10	3.6×10^{15}	−24.83	**0.00**	−21.21	2.73
		−23.37	7.39	−23.92	2.52
		−20.51	1.96	−20.75	2.53
		−23.70	2.25	−23.12	2.60
		−22.65	2.34	−19.63	2.72
12	4.7×10^{18}	−27.46	2.17	−28.16	2.55
		−27.37	1.30	−26.75	2.40
		−26.48	2.11	−26.78	1.98
		−25.96	1.84	−29.71	3.11
		−27.02	2.02	−28.47	**0.00**
14	6.1×10^{21}	−32.01	2.52	−33.38	2.47
		−33.21	2.65	−32.31	3.04
		−33.67	2.32	−31.27	2.67
		−32.61	3.88	−33.23	2.83
		−35.59	**0.00**	−34.29	3.16
16	7.9×10^{24}	−38.42	3.27	−36.29	2.96
		−39.78	2.87	−40.71	1.85
		−40.63	3.45	−43.55	**0.00**
		−42.62	2.67	−40.81	2.09
		−36.00	3.25	−41.94	2.73

[a] DMEs are calculated with respect to the lowest energy conformation found with one of the methods. See Table III for abbreviations.

The Ramachandran diagrams[26] for both amino acids reveal that, for the polyalanines, the allowed conformations are, from a steric point of view, restricted to about 12% of the (ϕ, ψ) space. For the polyglycines, about 60% of the (ϕ, ψ) space is accessible. As a consequence, many more low energy structures with similar energies exist, and are distributed over a wide range of the conformational space; that is, structures with similar energies may have large DMEs.

Conclusions

The SSV method presented is an efficient and easily implementable strategy for finding low energy conformations of polypeptides. It is capable of locating the global energy minimum of Met-enkephalin, a peptide often employed to test the efficiency of approaches for overcoming the multiple minima problem more frequently than MMC-

SA: On average, the global minimum is found 6.8 times more frequently with SSV than with MMC-SA within the same number of steps. On the other hand, one step (rotation of a dihedral angle, energy minimization, and acceptance or rejection of the new structure) consumes about 1.5 times more CPU time for SSV, because, to reject or accept a structure, one or two rotations (in both directions for the same bond) with equal probability are needed. Thus, one can estimate that SSV consumes, on average, only 20% of the CPU time required for MMC-SA to find the global minimum of Met-enkephalin. Moreover, comparison of SSV with a sophisticated method, conformational space annealing (CSA),[18] shows that the global minimum is found after approximately the same number of energy minimizations (i.e., steps) with the SSV method.

The simulations of homopolymers of different lengths show that SSV is again more successful in finding the lowest energy structures than MMC-SA. The energy differences of structures located by the two methods increase as the magnitude of the problem increases.

Finally, problems arising from the choice of optimal simulation temperature or cooling schedule, respectively, disappear, as the SSV method is temperature-independent.

References

1. S. J. Weiner, P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profeta, and P. Weiner, J. Am. Chem. Soc., **106**, 765 (1984).
2. W. F. van Gunsteren and H. J. C. Berendsen, *Groningen Molecular Simulation* (*GROMOS*) *Library Manual*, Biomos, Groningen, 1987.
3. A. Momany, R. F. McGuire, A. W. Burgess, and H. A. Scheraga, J. Phys. Chem., **79**, 2361 (1975).
4. G. Némethy, M. S. Pottle, and H. A. Scheraga, J. Phys. Chem., **87**, 1883 (1983).
5. M. Sippl, G. Némethy, and H. A. Scheraga, J. Phys. Chem., **88**, 6231 (1984).
6. G. Némethy, K. D. Gibson, K. A. Palmer, C. N. Yoon, G. Paterlini, A. Zagari, S. Rumsey, and H. A. Scheraga, J. Phys. Chem., **96**, 6472 (1992).
7. D. Voet and J. G. Voet, Biochemistry, John Wiley & Sons, 1990.
8. M. Karplus and G. A. Petsko, Nature, **347**, 631 (1990).
9. N. Metropolis, A. W. Rosenbluth, M. B. Rosenbluth, and A. H. Teller, J. Phys. Chem., **21**, 1087 (1953).
10. Z. Li and H. A. Scheraga, Proc. Natl. Acad. Sci. USA, **84**, 6611 (1987).
11. S. Kirkpatrick, C. D. Gelatt, M. P. Vecchi, Science, **220**, 671 (1983).
12. S. R. Wilson and W. Cui, Biopolymers, **29**, 225 (1990).
13. J. Ma and J. E. Straub, J. Chem. Phys., **101**, 533 (1994).
14. U. H. E. Hansmann and Y. Okamoto, J. Comput. Chem., **14**, 1333 (1993).
15. I. Andricioaiei and J. E. Straub, Phys. Rev., **E53**, 3055 (1996).
16. R. S. Judson, E. P. Jaeger, A. M. Treasurywala, and M. Peterson, J. Comput. Chem., **14**, 1407 (1993).
17. M. Vásquez, G. Némethy, and H. A. Scheraga, Chem. Rev., **94**, 2183 (1994).
18. J. Lee, H. A. Scheraga, and S. Rackovsky, J. Comput. Chem., **18**, 1222 (1997).
19. L. Piela, J. Kostrowicki, and H. A. Scheraga, J. Phys. Chem., **93**, 3339 (1989).
20. J. Kostrowicki and H. A. Scheraga, J. Phys. Chem., **96**, 7442 (1992).
21. P. J. Kraulis, G. M. Clore, M. Nigels, T. Jones, G. Petterson, J. Knowles, and A. M. Gronenborn, Biochemistry, **28**, 7241 (1989).
22. T. Schaumann, W. Braun, and Wütherich, Biopolymers, **29**, 679 (1990).
23. B. von Freyberg and W. Braun, J. Comput. Chem., **12**, 1065 (1991).
24. T. Solmajer and E. L. Mehler, Int. J. Quant. Chem., **44**, 291 (1992).
25. D. M. Gay, ACM Trans. Math. Software, **9**, 503 (1983).
26. D. Voet and J. G. Voet, Biochemistry, John Wiley & Sons, 1990.